Published in final edited form as:
Science. 2018 December 07; 362(6419): 1161–1164. doi:10.1126/science.aar6731.

Quantifying the contribution of recessive coding variation to developmental disorders

Hilary C. Martin[1,*], Wendy D. Jones[1,2], Rebecca McIntyre[1], Gabriela Sanchez-Andrade[1], Mark Sanderson[1], James D. Stephenson[1,3], Carla P. Jones[1], Juliet Handsaker[1], Giuseppe Gallone[1], Michaela Bruntraeger[1], Jeremy F. McRae[1], Elena Prigmore[1], Patrick Short[1], Mari Niemi[1], Joanna Kaplanis[1], Elizabeth J. Radford[1,4], Nadia Akawi[5], Meena Balasubramanian[6], John Dean[7], Rachel Horton[8], Alice Hulbert[9], Diana S. Johnson[6], Katie Johnson[10], Dhavendra Kumar[11], Sally Ann Lynch[12], Sarju G. Mehta[13], Jenny Morton[14], Michael J. Parker[15], Miranda Splitt[16], Peter D Turnpenny[17], Pradeep C. Vasudevan[18], Michael Wright[16], Andrew Bassett[1], Sebastian S. Gerety[1], Caroline F. Wright[19], David R. FitzPatrick[20], Helen V. Firth[1,13], Matthew E. Hurles[1], Jeffrey C. Barrett[1,*] on behalf of the DDD Study

[1]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, U.K [2]Great Ormond Street Hospital for Children, NHS Foundation Trust, Great Ormond Street Hospital, Great Ormond Street, London WC1N 3JH, UK [3]European Molecular Biology Laboratory–European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK [4]Department of Paediatrics, Cambridge University Hospitals NHS Foundation Trust, Cambridge, U.K [5]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, U.K [6]Sheffield Clinical Genetics Service, Sheffield Children's NHS Foundation Trust, OPD2, Northern General Hospital, Herries Rd, Sheffield, S5 7AU, U.K [7]Department of Genetics, Aberdeen Royal Infirmary, Aberdeen, U.K [8]Wessex Clinical Genetics Service, G Level, Princess Anne Hospital, Coxford Road, Southampton, SO16 5YA [9]Cheshire and Merseyside Clinical Genetic Service, Liverpool Women's NHS Foundation Trust, Crown Street, Liverpool, L8 7SS, U.K [10]Department of Clinical Genetics, City Hospital Campus, Hucknall Road, Nottingham, NG5 1PB, U.K [11]Institute of Cancer and Genetics, University Hospital of Wales, Cardiff, U.K [12]Temple Street Children's Hospital, Dublin, Ireland [13]Department of Clinical Genetics, Cambridge University Hospitals NHS Foundation Trust, Cambridge, U.K [14]Clinical Genetics Unit, Birmingham Women's Hospital, Edgbaston, Birmingham, B15 2TG, U.K [15]Sheffield Clinical Genetics Service, Sheffield Children's Hospital, Western Bank, Sheffield, S10 2TH, U.K [16]Northern Genetics Service, Newcastle upon Tyne Hospitals, NHS Foundation Trust [17]Clinical Genetics, Royal Devon & Exeter NHS Foundation Trust, Exeter, U.K [18]Department of Clinical

[*]Corresponding authors: hcm@sanger.ac.uk and jeff.barrett@genomicsplc.com.

Author contributions: Data analysis: H.C.M., J.F.M., J.H., P.S., N.A.; Clinical interpretation: W.D.J.; *EIF3F* experiments: R.M., C.P.J.., M.B.; Mouse phenotyping: G.S.-A., M.S.; Protein structure modelling: J.D.S.; Data processing: G.G., M.N., J.K., C.F.W., E.R.; Experimental validation: E.P.; Patient recruitment: M. Balasubramanian, J.D., R.H., A.H., D.S.J., K.J., D.K, S.A.L., S.G.M., J.M., M.J.P., M.S., P.D.T., P.C.V., M.W.; Experimental and analytical supervision: A.B., S.S.G., C.F.W., D.R.F., H.V.F., M.E.H., J.C.B.; Writing: H.C.M., W.D.J., R.M., G.S.-A., J.S., M.B., A.B., J.H., M.E.H., J.C.B.

Competing interests: M.E.H. is a co-founder of, consultant to, and holds shares in, Congenica Ltd, a genetics diagnostic company.

Data and materials availability: Exome sequencing and phenotype data are accessible via the European Genome-phenome Archive (EGA) (Datafreeze 2016-10-03) (https://www.ebi.ac.uk/ega/studies/EGAS00001000775).

Genetics, University Hospitals of Leicester NHS Trust, Leicester Royal Infirmary, Leicester, LE1 5WW [19]University of Exeter Medical School, Institute of Biomedical and Clinical Science, RILD, Royal Devon & Exeter Hospital, Barrack Road, Exeter, EX2 5DW, U.K [20]MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, U.K

Abstract

We estimated the genome-wide contribution of recessive coding variation from 6,040 families from the Deciphering Developmental Disorders study. The proportion of cases attributable to recessive coding variants was 3.6% in patients of European ancestry, compared to 50% explained by *de novo* coding mutations. It was higher (31%) in patients with Pakistani ancestry, due to elevated autozygosity. Half of this recessive burden is attributable to known genes. We identified two genes not previously associated with recessive developmental disorders, *KDM5B* and *EIF3F*, and functionally validated them with mouse and cellular models. Our results suggest that recessive coding variants account for a small fraction of currently undiagnosed non-consanguineous individuals, and that the role of noncoding variants, incomplete penetrance, and polygenic mechanisms need further exploration.

Large-scale sequencing studies of phenotypically heterogeneous rare disease patients can discover new disease genes (1–3) and characterise the genetic architecture of such disorders. In the Deciphering Developmental Disorders (DDD) study, we previously estimated the fraction of patients with a causal *de novo* coding mutation in both known and as-yet-undiscovered disease genes to be 40-45% (4), and here we extend this approach to recessive variants. It has been posited that there are thousands of as-yet-undiscovered recessive intellectual disability (ID) genes (5, 6), which could imply that recessive variants explain a large fraction of undiagnosed rare disease cases. However, attempts to estimate the prevalence of recessive disorders have been restricted to known disorders (7) or known pathogenic alleles (8). Here, we quantify the total autosomal recessive coding burden using a robust and unbiased statistical framework in 6,040 exome-sequenced DDD trios from the British Isles. Our approach provides a better-calibrated estimate of the exome-wide burden of recessive disease than previously published methods (3, 9).

We analysed 5,684 European and 356 Pakistani probands (EABI, PABI - European or Pakistani Ancestry from the British Isles; Fig. S1, S2) with developmental disorders (DDs). The clinical features are heterogeneous and representative of genetically undiagnosed DD patients from British and Irish clinical genetics services: 88% have an abnormality of the nervous system, and 88% have multiple affected organ systems (Fig. 1, Fig. S3, Table S1). Clinical features are largely similar between EABI and PABI (Fig. 1, Table S1).

To assess the genome-wide recessive burden, we compared the number of rare (minor allele frequency, MAF, <1%) biallelic genotypes observed in our cohort to the number expected by chance (10). We used the phased haplotypes from unaffected DDD parents to estimate the expected number of biallelic genotypes. Reassuringly, the number of observed biallelic synonymous genotypes matched the expectation (Fig. S4). We observed no significant burden of biallelic genotypes of any consequence class in 1,389 probands with a likely

diagnostic *de novo*, inherited dominant or X-linked variant. We therefore evaluated the recessive coding burden in the remaining 4,318 EABI and 333 PABI probands. This "undiagnosed" cohort were more likely to have a recessive cause because they did not have a likely dominant or X-linked diagnosis (11), had at least one affected sibling, or >2% autozygosity (Fig. 2A). As expected due to their higher autozygosity (Fig. S5), PABI individuals had more rare biallelic genotypes than EABI individuals (Fig. 2A); 92% of these were homozygous (rather than compound heterozygous), versus only 28% for the EABI samples. We observed a significant enrichment of biallelic loss-of-function (LoF) genotypes in both undiagnosed ancestry groups (Poisson $p=3.5x10^{-5}$ in EABI, $p=9.7x10^{-7}$ in PABI), and, in the EABI group, a nominally significant enrichment of biallelic damaging missense genotypes (p=0.025) and a significant enrichment of compound heterozygous LoF/damaging missense genotypes ($p=6x10^{-7}$) (Fig. 2A).

Amongst the 4,651 EABI+PABI undiagnosed probands, a set of 903 clinically-curated DD-associated recessive genes showed a higher recessive burden (Fig. S6; 1.7-fold; Poisson $p=6\times10^{-18}$) than average (1.1-fold for all genes). Indeed, 48% of the observed excess of biallelic genotypes lay in these known genes. By contrast, we did not observe any recessive burden in 243 DD-associated genes with a dominant LoF mechanism, nor in any gene sets tested in the 1,389 diagnosed probands (Poisson p>0.05).

We developed a method to estimate the proportion of probands with a causal variant in a particular genotype class (10) in either known and as-yet-undiscovered genes. Unlike our previously published approach (4), this method accounts for the fact that some fraction of the variants expected by chance are actually causal (Fig. S7). We estimated that 3.6% (~205) of the 5,684 EABI probands have a recessive coding diagnosis, compared to 49.9% (~2836) with a *de novo* coding diagnosis. Recessive coding genotypes explain 30.9% (~110) of the 356 PABI individuals, compared to 29.8% (~106) for *de novos*. The contribution from recessive variants was higher in EABI probands with affected siblings than those without (12.0% of 117 versus 3.2% of 5,098), and highest in PABI probands with high autozygosity (47.1% of 241) (Fig. 2B; Table S2). In contrast, it did not differ between 115 PABI probands with low autozygosity and all 5,684 EABI probands.

We caution that the PABI results may be less reliable due to modest sample size (note the wide confidence intervals in Table S2), exacerbated by consistent overestimation of rare variant frequencies in our limited sample of parents. Reassuringly, our estimated recessive contribution in PABI is close to the 31.5% reported in Kuwait (12), which has a similar level of consanguinity (13). Our results are consistent with previous reports of a low fraction of recessive diagnoses in European cohorts (3, 11, 14), but unlike those studies, our estimates further show that the recessive contribution in as-yet-undiscovered genes is also small. While it has been hypothesised that there are thousands of undiscovered recessive DD-associated genes (5, 6), our analyses suggest that the cumulative impact of these discoveries on diagnostic yield will be modest in non-consanguineous populations.

We next tested each gene for an excess of biallelic genotypes in the undiagnosed probands (Table S3) (10). Three genes passed stringent Bonferroni correction ($p<3.4\times10^{-7}$) (10), *THOC6* (previously reported (15)), *EIF3F*, and *KDM5B*. Thirteen additional genes had

$p<10^{-4}$ (Table S4), of which eleven are known recessive DD-associated genes, and known genes were enriched for lower p-values (Fig. S8).

We observed five probands with an identical homozygous missense variant in *EIF3F* (binomial $p=1.2\times10^{-10}$) (ENSP00000310040.4:p.Phe232Val), plus four additional homozygous probands who had been excluded from our discovery analysis for various reasons (Table S5). The variant (rs141976414) has a frequency of 0.12% in non-Finnish Europeans (one of the most common protein-altering variants in the gene), and no homozygotes were observed in gnomAD (http://gnomad.broadinstitute.org/).

All nine individuals homozygous for Phe232Val had intellectual disability (ID) and a subset also had seizures (6/9), behavioral difficulties (3/9) and sensorineural hearing loss (3/9) (Table S5). There was no obvious distinctive facial appearance (Fig. S9). *EIF3F* encodes a subunit of the mammalian eIF3 (eukaryotic initiation factor) complex, which negatively regulates translation. The genes encoding eIF2B subunits have been implicated in severe autosomal recessive neurodegenerative disorders (16). We edited iPSC lines with CRISPR-Cas9 to be heterozygous or homozygous for the Phe232Val variant, and Western blots showed that EIF3F protein levels were ~27% lower in homozygous cells relative to heterozygous and wild-type cells (Fig. S10), which may be due to reduced protein stability (Fig. S11). The Phe232Val variant significantly reduced translation rate (Fig. 3A, Fig. S12). Proliferation rates were also reduced in the homozygous but not heterozygous cells (Fig. 3B, Fig. S13), although the viability of the cells was unchanged (Fig. S14).

Another recessive gene we identified was *KDM5B* (binomial $p=1.1\times10^{-7}$) (Fig. 4), encoding a histone H3K4 demethylase. Three probands had biallelic LoFs passing our filters, and a fourth was compound heterozygous for a splice site variant and a large gene-disrupting deletion. Several of these patients were recently reported with less compelling statistical evidence (17). Interestingly, *KDM5B* is also enriched for *de novo* mutations in our cohort (4) (binomial $p=5.1\times10^{-7}$). We saw nominally significant over-transmission of LoFs from the mostly unaffected parents (p=0.002, transmission-disequilibrium test; Table S6), but no parent-of-origin bias. Theoretically, all the *KDM5B* LoFs observed in probands might be acting recessively and heterozygous probands may have a second (missed) coding or regulatory hit or modifying epimutation. However, we found no evidence supporting this (see (10); Fig. S15, S16), nor of potentially modifying coding variants in likely interactor genes, nor that some LoFs avoid nonsense-mediated decay (Fig. 4B). Genome-wide levels of DNA methylation in whole blood did not differ between probands with different types of *KDM5B* mutations or between these and controls (Fig. S17).

These lines of evidence, along with previous observations of *KDM5B de novos* in both autism patients and unaffected siblings (18), suggest that heterozygous LoFs in *KDM5B* are pathogenic with incomplete penetrance, while homozygous LoFs are likely fully penetrant. Several microdeletions (19) and LoFs in other dominant ID genes are incompletely penetrant (20). Other H3K4 methylases and demethylases also cause neurodevelopmental disorders (21). *KDM5B* is atypical; the others are mostly dominant (21), typically with pLI scores >0.99 and very low pRec scores, whereas *KDM5B* has $pLI=5\times10^{-5}$ and pRec>0.999 (22).

KDM5B is the only gene that showed significant enrichment for both biallelic variants and *de novo* mutations in our study. We saw significant enrichment of *de novo* missense (373 observed versus 305 expected; ratio=1.25, upper-tailed Poisson $p=1x10^{-4}$) but not *de novo* LoF mutations across all known recessive DD genes (excluding those known to also show dominant inheritance). One hypothesis is that the *de novo* missense mutations are acting as a "second hit" on the opposite haplotype from an inherited variant in the same gene. However, we saw only two instances of this in the cohort, and besides, if it were driving the signal, we would expect to see a burden of *de novo* LoFs in recessive genes too, which we do not. A better explanation is that recessive DD genes are also enriched for dominant activating mutations. There are known examples of this; e.g. in *NALCN* (23, 24) and *MAB21L2* (25), heterozygous missense variants are activating or dominant-negative, whereas the biallelic mechanism is loss-of-function. In contrast, the six *de novo* LoFs in *KDM5B* suggest it follows a different pattern. Of the twenty-one recessive genes with nominally significant *de novo* missense enrichment in our data, only one showed evidence of mutation clustering using our previously published method (1) (*CTC1*; p=0.03), which could suggest an activating/dominant-negative mechanism. Larger sample sizes will be needed to establish which of these genes also act dominantly, and by which mechanism.

All four individuals with biallelic *KDM5B* variants have ID, variable congenital abnormalities (Table S7) and a distinctive facial appearance (Fig. S18). Other than ID, there were no consistent phenotypes or distinctive features shared between the biallelic and monoallelic individuals, or within the monoallelic group (Table S7).

We created a mouse loss-of-function model for *Kdm5b*. Heterozygous knockout mice appear normal and fertile, while homozygous *Kdm5b*-null mice are subviable (44% of expected, from heterozygous in-crosses). This partially penetrant lethality, in addition to a fully penetrant vertebral patterning defect (Fig. S19), is consistent with previously published work (26). We additionally identified numerous behavioral abnormalities in homozygous *Kdm5b*-null mice: increased anxiety, less sociability, and reduced long-term memory compared to wild-types (Fig. 4).

We have quantified the contribution of recessive coding variants in both known and as-yet-undiscovered genes to a large UK cohort of DD patients, and found that overall they explain a small fraction. Our methodology allowed us to carry out an unbiased burden analysis not possible with previous methods (Fig. S4). We identified two new recessive DD genes that are less likely to be found by typical studies because they result in heterogeneous and nonspecific phenotypes, and presented strong functional evidence supporting their pathogenicity.

Our results can be used to improve recurrence risk estimates for undiagnosed families with a particular ancestry and pattern of inheritance. Extrapolating our results more widely requires some care: our study is slightly depleted of recessive diagnoses since some recessive DDs (e.g. metabolic disorders) are relatively easily diagnosed through current clinical practice in the UK and less likely to have been recruited. Furthermore, country-specific diagnostic practices and levels of consanguinity may make the exact estimates less applicable outside the UK.

Europe PMC Funders Author Manuscripts

Overall, we estimated that identifying all recessive DD genes would allow us to diagnose 5.2% of the EABI+PABI subset of DDD, whereas identifying all dominant DD genes would yield diagnoses for 48.6%. The high proportion of unexplained patients even amongst those with affected siblings or high consanguinity suggests that future studies should investigate a wide range of modes of inheritance including oligogenic and polygenic inheritance as well as noncoding recessive variants.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the DDD families, the Sanger Human Genome Informatics team, P. Danacek for help with bcftools/roh, K. de Lange for help with figures, K. Samocha for mutability estimates, J. Matte and G. Turner for help with experiments, and A. Sakar for patient review.

Families gave informed consent to participate, and the study was approved by the UK Research Ethics Committee (10/H0305/83, granted by the Cambridge South Research Ethics Committee and GEN/284/12, granted by the Republic of Ireland Research Ethics Committee).

Funding: The DDD study presents independent research commissioned by the Health Innovation Challenge Fund [grant number HICF-1009-003]. See Supplement for details.

References and Notes

1. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. Nature. 2015; 519:223–228. [PubMed: 25533962]
2. Yuen RKC, et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. Nat Neurosci. 2017; 20
3. Jin SC, et al. Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. Nat Genet. 2017; doi: 10.1038/ng.3970
4. Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. Nature. 2017; 542:433–438. [PubMed: 28135719]
5. Ropers HH. Genetics of early onset cognitive impairment. Annu Rev Genomics Hum Genet. 2010; 11:161–187. [PubMed: 20822471]
6. Vissers LELM, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. Nat Rev Genet. 2016; 17:9–18. [PubMed: 26503795]
7. Baird PA, Anderson TW, Newcombe HB, Lowry RB. Genetic disorders in children and young adults: a population study. Am J Hum Genet. 1988; 42:677–693. [PubMed: 3358420]
8. Schrodi SJ, et al. Prevalence estimation for monogenic autosomal recessive diseases using population-based genetic data. Hum Genet. 2015; 134:659–669. [PubMed: 25893794]
9. Akawi N, et al. Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families. Nat Genet. 2015; 47:1363–1369. [PubMed: 26437029]
10. Materials and methods are available as supplementary materials at the Science website.
11. Wright CF, et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. Lancet. 2015; 385:1305–1314. [PubMed: 25529582]
12. Teebi AS. Autosomal recessive disorders among Arabs: an overview from Kuwait. J Med Genet. 1994; 31:224–233. [PubMed: 8014972]
13. Tadmouri GO, et al. Consanguinity and reproductive health among Arabs. Reprod Health. 2009; 6:17. [PubMed: 19811666]

14. Gilissen C, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature. 2014; 511:344–347. [PubMed: 24896178]
15. Beaulieu CL, et al. Intellectual disability associated with a homozygous missense mutation in THOC6. Orphanet J Rare Dis. 2013; 8:62. [PubMed: 23621916]
16. Fogli A, Boespflug-Tanguy O. The large spectrum of eIF2B-related diseases. Biochem Soc Trans. 2006; 34:22–29. [PubMed: 16246171]
17. Faundes V, et al. Histone Lysine Methylases and Demethylases in the Landscape of Human Developmental Disorders. Am J Hum Genet. 2018; 102:175–187. [PubMed: 29276005]
18. Iossifov I, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014; 515:216–221. [PubMed: 25363768]
19. Carvill GL, Mefford HC. Microdeletion syndromes. Curr Opin Genet Dev. 2013; 23:232–239. [PubMed: 23664828]
20. Ropers HH, Wienker T. Penetrance of pathogenic mutations in haploinsufficient genes for intellectual disability and related disorders. Eur J Med Genet. 2015; 58:715–718. [PubMed: 26506440]
21. Vallianatos CN, Iwase S. Disrupted intricacy of histone H3K4 methylation in neurodevelopmental disorders. Epigenomics. 2015; 7:503–519. [PubMed: 26077434]
22. Lek M, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016; 536:285–291. [PubMed: 27535533]
23. Chong JX, et al. De novo mutations in NALCN cause a syndrome characterized by congenital contractures of the limbs and face, hypotonia, and developmental delay. Am J Hum Genet. 2015; 96:462–473. [PubMed: 25683120]
24. Al-Sayed MD, et al. Mutations in NALCN cause an autosomal-recessive syndrome with severe hypotonia, speech impairment, and cognitive delay. Am J Hum Genet. 2013; 93:721–726. [PubMed: 24075186]
25. Rainger J, et al. Monoallelic and biallelic mutations in MAB21L2 cause a spectrum of major eye malformations. Am J Hum Genet. 2014; 94:915–923. [PubMed: 24906020]
26. Albert M, et al. The histone demethylase Jarid1b ensures faithful mouse development by protecting developmental genes from aberrant H3K4me3. PLoS Genet. 2013; 9:e1003461. [PubMed: 23637629]
27. Kohler S, et al. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. Am J Hum Genet. 2009; 85:457–464. [PubMed: 19800049]
28. Bragin E, et al. DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. Nucleic Acids Res. 2014; 42:D993–D1000. [PubMed: 24150940]
29. Sheridan E, et al. Risk factors for congenital anomaly in a multiethnic birth cohort: an analysis of the Born in Bradford study. Lancet. 2013; 382:1350–1359. [PubMed: 23830354]
30. McLaren W, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016; 17:122. [PubMed: 27268795]
31. 1000 Genomes Project Consortium. et al. A global reference for human genetic variation. Nature. 2015; 526:68–74. [PubMed: 26432245]
32. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–909. [PubMed: 16862161]
33. Narasimhan V, et al. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics. 2016; 32:1749–1751. [PubMed: 26826718]
34. Conomos MP, Reiner AP, Weir BS, Thornton TA. Model-free Estimation of Recent Genetic Relatedness. Am J Hum Genet. 2016; 98:127–148. [PubMed: 26748516]
35. Samocha KE, et al. A framework for the interpretation of de novo mutation in human disease. Nat Genet. 2014; 46:944–950. [PubMed: 25086666]
36. Krissinel E, Henrick K. Multiple alignment of protein structures in three dimensions. CompLife. 2005; 3695:67–78.

37. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr. 2004; 60:2256–2268. [PubMed: 15572779]
38. Hubbard, SJ, Thornton, JM. NACCESS-Computer Program. Department of Biochemistry and Molecular Biology, University College; London: 1993.
39. Valdar WSJ. Scoring residue conservation. Proteins. 2002; 48:227–241. [PubMed: 12112692]
40. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004; 14:1188–1190. [PubMed: 15173120]
41. Kilpinen H, et al. Common genetic variation drives molecular heterogeneity in human iPSCs. Nature. 2017; 546:370–375. [PubMed: 28489815]
42. Knapp M. The transmission/disequilibrium test and parental-genotype reconstruction: the reconstruction-combined transmission/disequilibrium test. Am J Hum Genet. 1999; 64:861–870. [PubMed: 10053021]
43. Szklarczyk D, et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Res. 2017; 45:D362–D368. [PubMed: 27924014]
44. Klein BJ, et al. The histone-H3K4-specific demethylase KDM5B binds to its substrate and product through distinct PHD fingers. Cell Rep. 2014; 6:325–335. [PubMed: 24412361]
45. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013; 10:5–6. [PubMed: 23269371]
46. Mali P, et al. RNA-guided human genome engineering via Cas9. Science. 2013; 339:823–826. [PubMed: 23287722]
47. Gapp K, et al. Potential of Environmental Enrichment to Prevent Transgenerational Effects of Paternal Trauma. Neuropsychopharmacology. 2016; 41:2749–2758. [PubMed: 27277118]
48. Dias C, et al. BCL11A Haploinsufficiency Causes an Intellectual Disability Syndrome and Dysregulates Transcription. Am J Hum Genet. 2016; 99:253–274. [PubMed: 27453576]
49. Csibi A, et al. The translation regulatory subunit eIF3f controls the kinase-dependent mTOR signaling required for muscle differentiation and hypertrophy in mouse. PLoS One. 2010; 5:e8994. [PubMed: 20126553]
50. Frank MC, Braginsky M, Yurovsky D, Marchman VA. Wordbank: an open repository for developmental vocabulary data. J Child Lang. 2017; 44:677–694. [PubMed: 27189114]
51. Smith ME, Lecker G, Dunlap JW, Cureton EE. The Effects of Race, Sex, and Environment on the Age at Which Children Walk. The Pedagogical Seminary and Journal of Genetic Psychology. 1930; 38:489–498.
52. Yamamoto GL, et al. Rare variants in SOS2 and LZTR1 are associated with Noonan syndrome. J Med Genet. 2015; 52:413–421. [PubMed: 25795793]
53. Akawi NA, Al-Jasmi F, Al-Shamsi AM, Ali BR, Al-Gazali L. LINS, a modulator of the WNT signaling pathway, is involved in human cognition. Orphanet J Rare Dis. 2013; 8:87. [PubMed: 23773660]
54. Najmabadi H, et al. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature. 2011; 478:57–63. [PubMed: 21937992]

Europe PMC Funders Author Manuscripts

One Sentence Summary

Recessive coding variants explain a low fraction of undiagnosed developmental disorder patients.

Europe PMC Funders Author Manuscripts

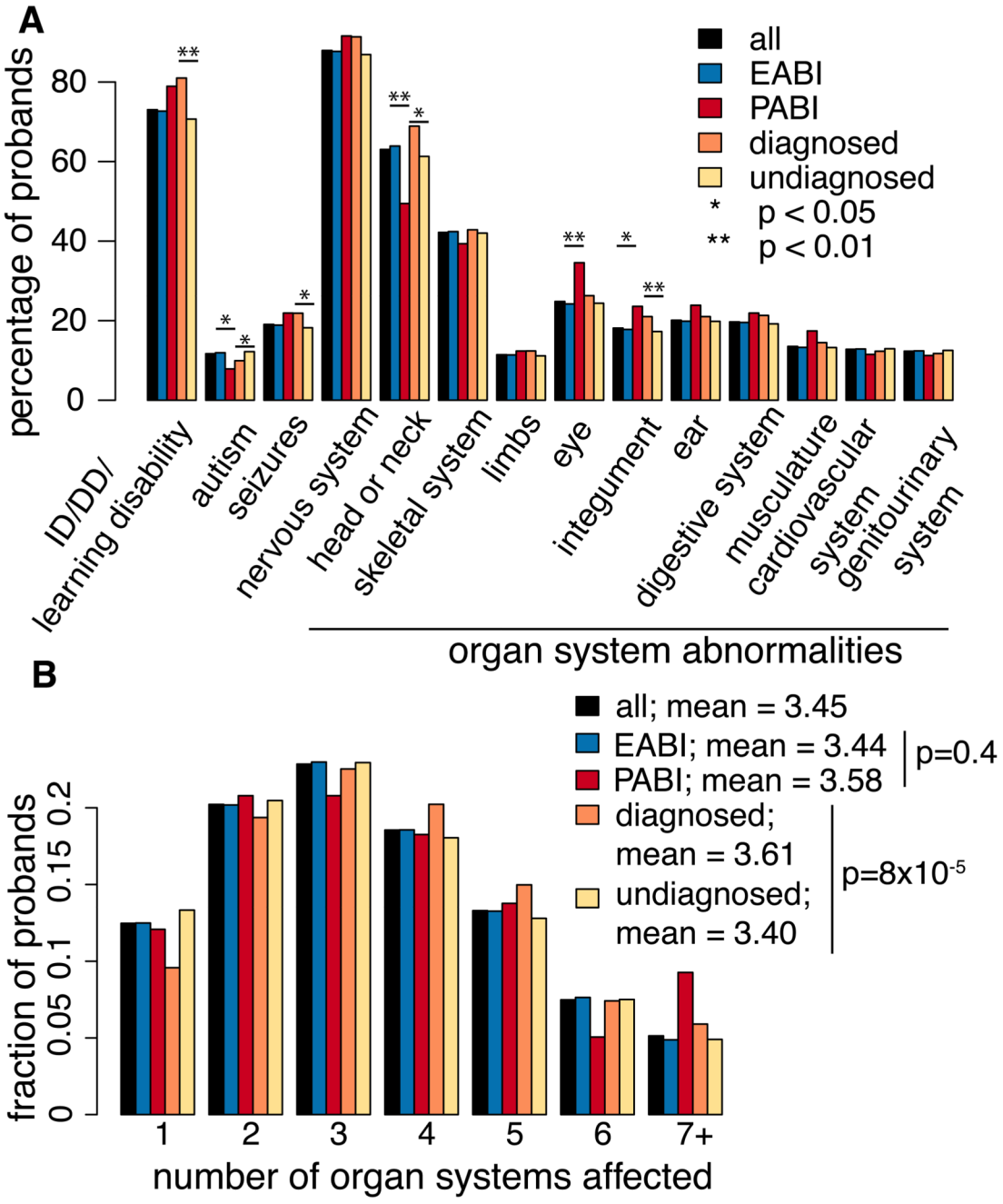

Fig. 1.

Clinical features of DDD probands analysed here. Proportion of probands in different groups with clinical features indicated, extracted from HPO terms. Asterisks indicate nominally significant differences between indicated groups (Fisher's exact test).

Europe PMC Funders Author Manuscripts

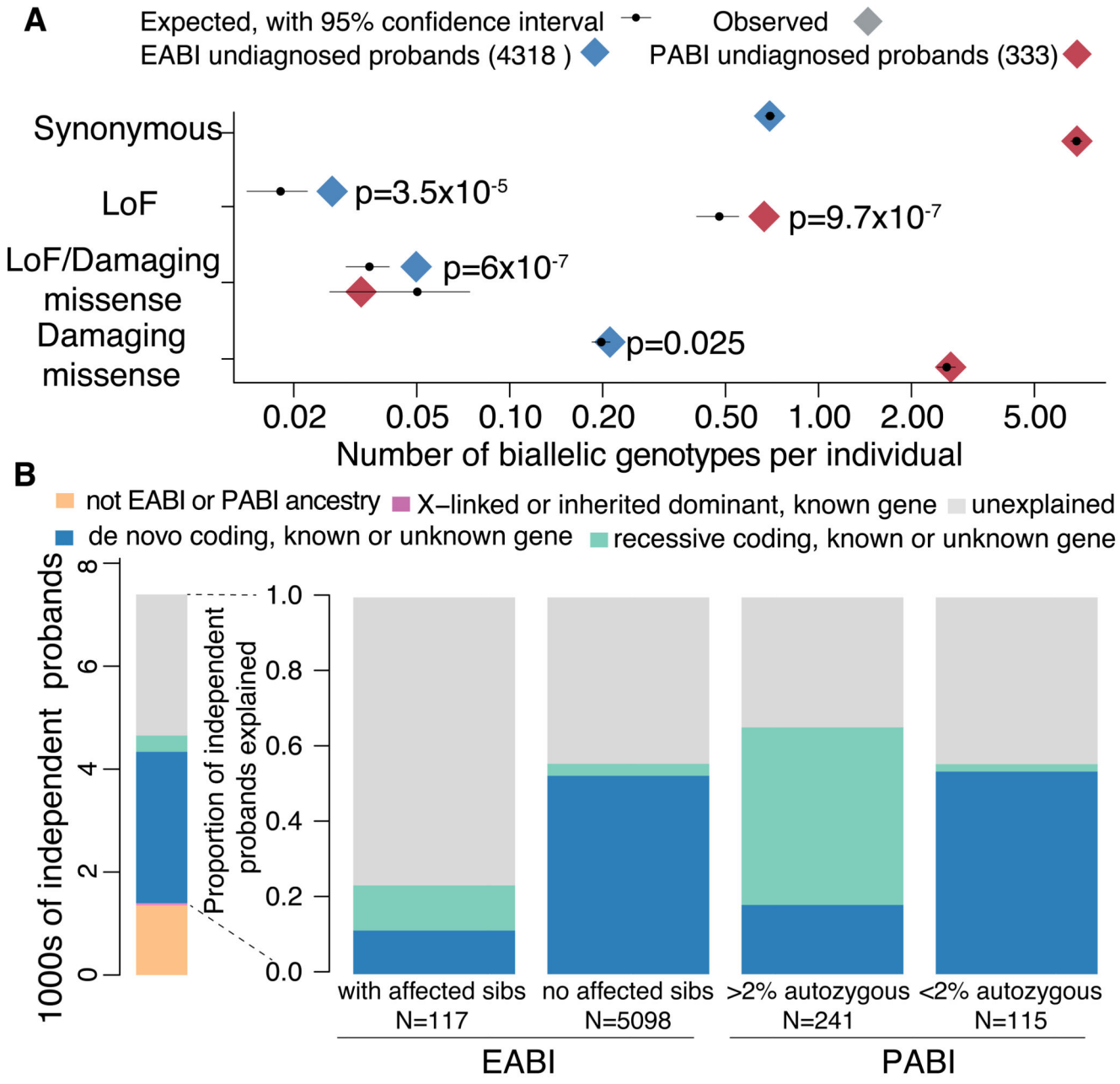

Fig. 2.

Contribution of recessive coding variants to genetic architecture in this study. (A) Number of observed and expected biallelic genotypes per individual across all genes. Nominally significant p-values from a Poisson test of enrichment are shown. (B) Left: number of probands grouped by diagnostic category. The inherited dominant and X-linked diagnoses (narrow pink bar) include only those in known genes, whereas the proportion of probands with *de novo* and recessive coding diagnoses was inferred as described in (10), including those in as-yet-undiscovered genes. Right: the proportion of probands in various patient subsets inferred to have diagnostic variants in the indicated classes.

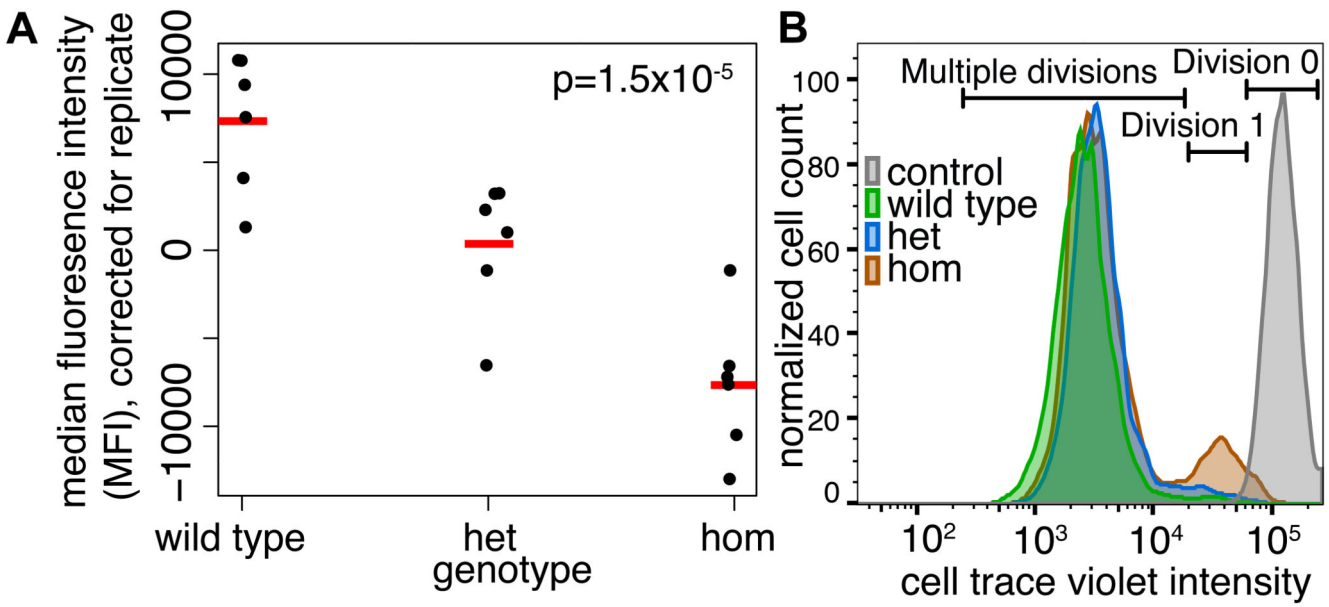

Fig. 3.

Functional consequences of the pathogenic *EIF3F* recessive missense variant. A) The Phe232Val variant impairs translation. Plot shows median fluorescence intensity (MFI) in iPSC lines heterozygous or homozygous for or without the Phe232Val variant (correcting for replicate effects), measured using a Click-iT protein synthesis assay (10). MFI correlates with methionine analogue incorporation in nascent proteins. The p-value indicates a non-zero effect of genotype from a linear regression of MFI on genotype and replicate. Red lines: means. B) The Phe232Val variant impairs iPSC proliferation in the homozygous but not heterozygous form. Results from a cell trace violet (CTV) proliferation assay, in which CTV concentration reduces on each division. The population of cells that have been through zero, one or multiple divisions is labelled.

Europe PMC Funders Author Manuscripts

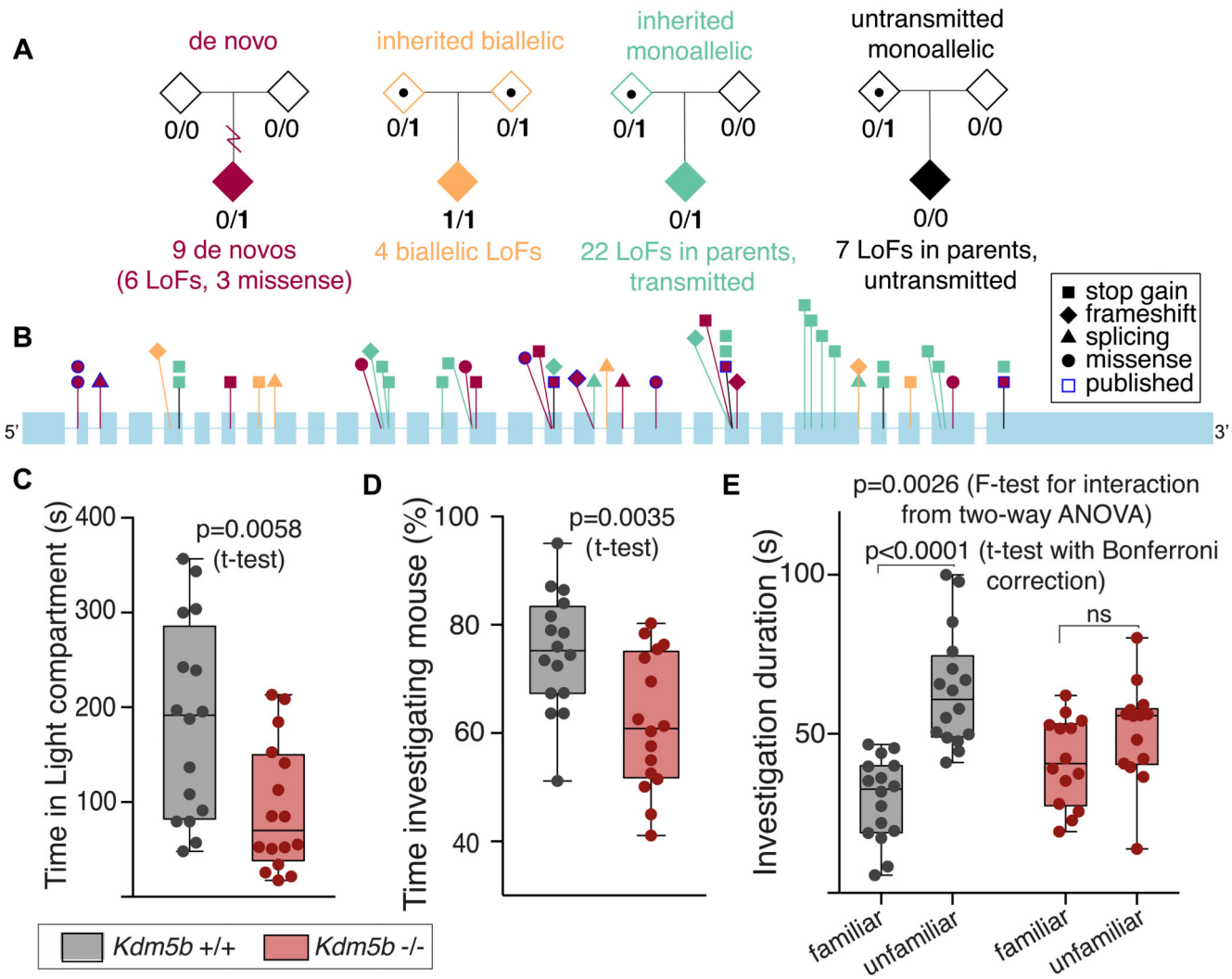

Fig. 4.

KDM5B is a recessive DD gene in which heterozygous LoFs are incompletely penetrant. A) Summary of damaging variants found in *KDM5B*. B) Positions of likely damaging variants found in this and previous studies in *KDM5B* (ENST00000367264.2; introns not to scale), omitting two large deletions. Colors correspond to those shown in (A). There are no differences in the spatial distribution of LoFs by inheritance mode, nor in their likelihood of escaping nonsense-mediated decay by alternative splicing in GTex (https://gtexportal.org/home/). C-E) Behavioral defects of homozygous *Kdm5b*-null versus wild-type mice (n=14-16). C) Knockout mice displayed increased anxiety, spending significantly less time in the light compartment of the Light-Dark box. D) Reduced sociability, in the three-chamber sociability test. Knockout mice spent less time investigating a novel mouse. E) 24h memory impairment. While wild-type mice preferentially investigated an unfamiliar mouse over a familiar one, homozygous knockout mice showed no discrimination.

Europe PMC Funders Author Manuscripts